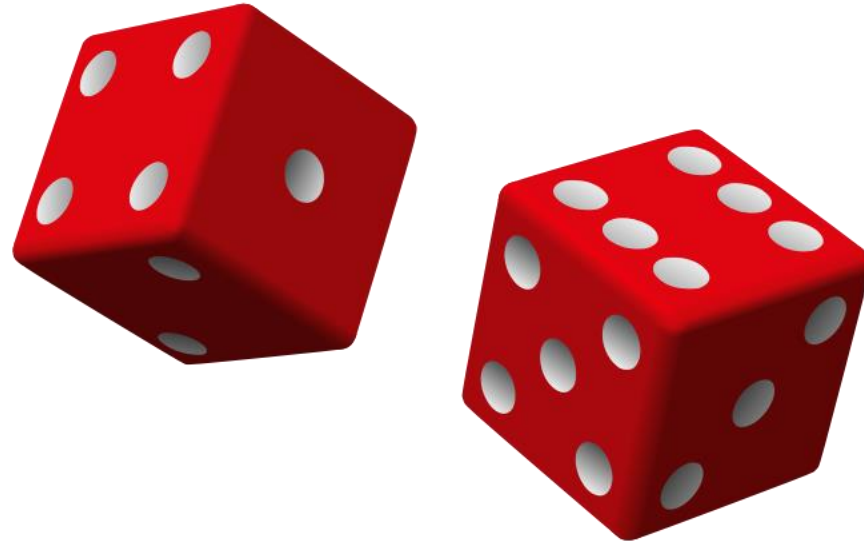


CS 6511: Artificial Intelligence

Probability



Amrinder Arora

The George Washington University

[Original version of these slides was created by Dan Klein and Pieter Abbeel for Intro to AI at UC Berkeley. <http://ai.berkeley.edu>]

Outline

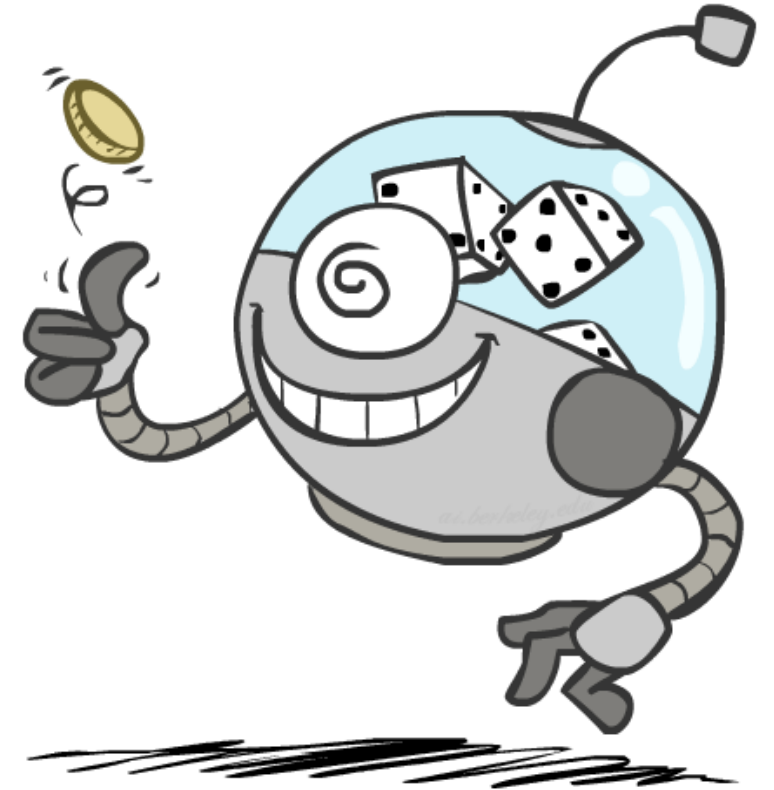
- New kinds of problems
 - Involve uncertainty
 - Variables are related to each other
 - Uncertainty and Random Variables
- Probability Distribution
 - Table
 - Joint Distribution
 - Marginal Distribution
 - Conditional Probability
 - Conditional Distribution
 - Normalization Trick
- Probabilistic Inference
 - Inference by Enumeration
 - Product Rule
 - Chain Rule
 - Bayes Rule
- Different Kinds of Probabilistic Models

Uncertainty

- General situation:
 - **Observed variables (evidence):** Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
 - **Unobserved variables:** Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
 - **Model:** Agent knows something about how the known variables relate to the unknown variables
- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge

Random Variables

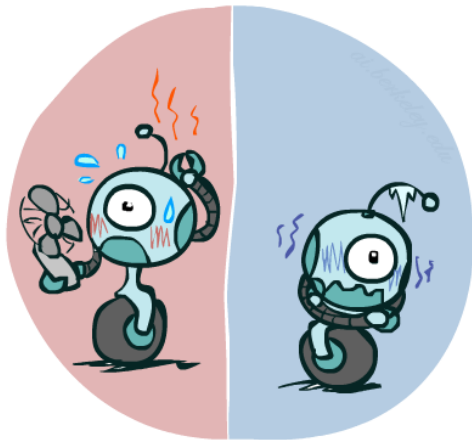
- A random variable is some aspect of the world about which we (may) have uncertainty
 - R = Is it raining?
 - T = Is it hot or cold?
 - D = How long will it take to drive to work?
 - L = Where is the ghost?
- We denote random variables with capital letters
- Like variables in a CSP, random variables have domains
 - R in {true, false} (often write as {+r, -r})
 - T in {hot, cold}
 - D in $[0, \infty)$
 - L in possible locations, maybe $\{(0,0), (0,1), \dots\}$



Probability Distributions

- Associate a probability with each domain value

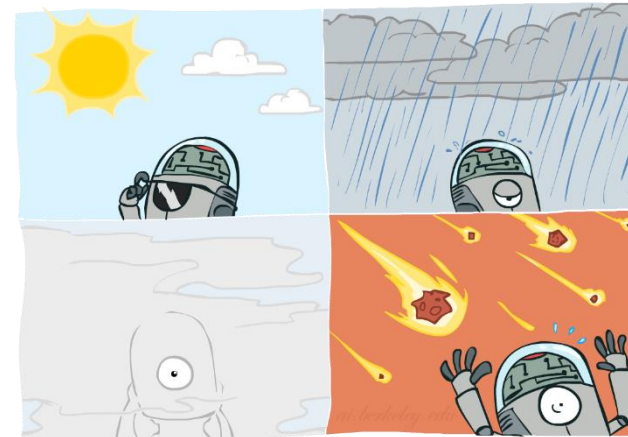
- Temperature:



$P(T)$

T	P
hot	0.5
cold	0.5

- Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Probability Distributions

- Unobserved random variables have distributions

$$P(T)$$

T	P
hot	0.5
cold	0.5

$$P(W)$$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Shorthand notation:

$$P(\textit{hot}) = P(T = \textit{hot}),$$

$$P(\textit{cold}) = P(T = \textit{cold}),$$

$$P(\textit{rain}) = P(W = \textit{rain}),$$

...

OK if all domain entries are unique

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number

$$P(W = \textit{rain}) = 0.1$$

- Must have: $\forall x \ P(X = x) \geq 0$ and $\sum_x P(X = x) = 1$

Joint Distributions

- A *joint distribution* over a set of random variables: X_1, X_2, \dots, X_n specifies a real number for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

$$P(x_1, x_2, \dots, x_n)$$

- Must obey: $P(x_1, x_2, \dots, x_n) \geq 0$

$$\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_n) = 1$$

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

- Size of distribution if n variables with domain sizes d?
 - For all but the smallest distributions, impractical to write out!

- nd
- $d! N$
- d^n
- n^d

X	Y	Z		
1	1	1		
1	1	2		
1	1	3		
1	2	1		
1	2	2		
1	2	3		
1	3	1		
1	3			

Events

- An *event* is a set E of outcomes

$$P(E) = \sum_{(x_1 \dots x_n) \in E} P(x_1 \dots x_n)$$

- From a joint distribution, we can calculate the probability of any event
 - Probability that it's hot AND sunny?
 - Probability that it's hot?
 - Probability that it's hot OR sunny?
- Typically, the events we care about are *partial assignments*, like $P(T=\text{hot})$

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Quiz: Events

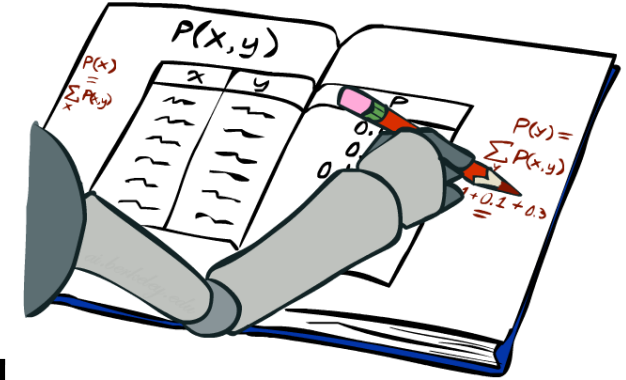
- $P(+x, +y)$?
- $P(+x)$?
- $P(-y \text{ OR } +x)$?

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding
 - “marginalization”: treating something as insignificant



$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



$$P(t) = \sum_s P(t, s)$$

$P(T)$

T	P
hot	0.5
cold	0.5



$$P(s) = \sum_t P(t, s)$$

$P(W)$

W	P
sun	0.6
rain	0.4

Marginalization

- In general terms, going from n variables to m ($< n$) variables

Quiz: Marginal Distributions

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1



$$P(x) = \sum_y P(x, y)$$



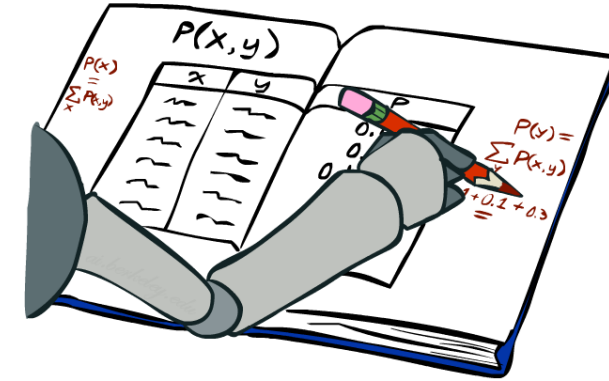
$$P(y) = \sum_x P(x, y)$$

$P(X)$

X	P
+x	
-x	

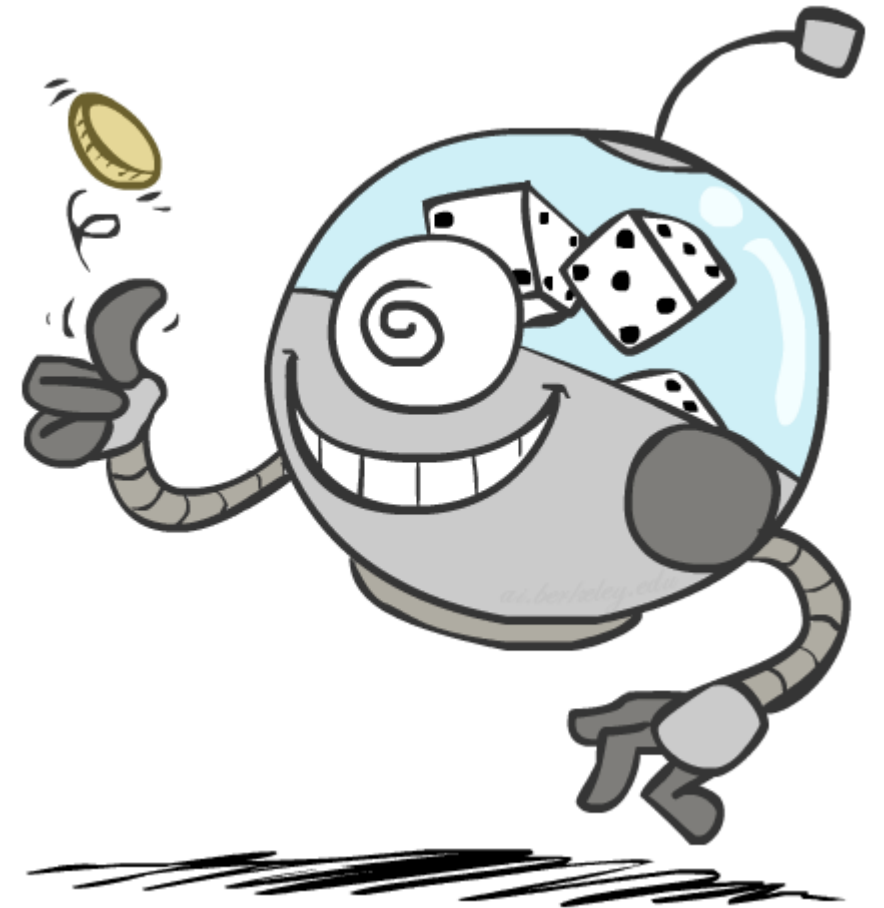
$P(Y)$

Y	P
+y	
-y	



Where we are...

- Probability
 - Random Variables
 - Joint and Marginal Distributions
 - Conditional Distribution
 - Product Rule, Chain Rule, Bayes' Rule
 - Inference
 - Independence



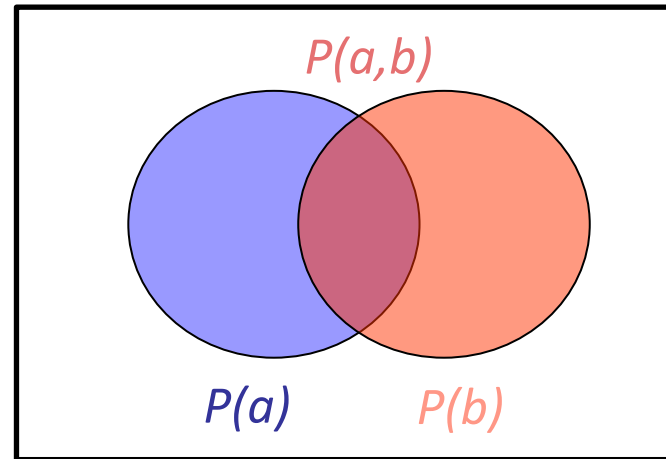
Conditional Probabilities

- A simple relation between joint and marginal probabilities

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



$$P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4$$

$$\begin{aligned} &= P(W = s, T = c) + P(W = r, T = c) \\ &= 0.2 + 0.3 = 0.5 \end{aligned}$$

Linda's Paradox

- *Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*
- Which is more probable?
 - Linda is a bank teller.
 - Linda is a bank teller and is active in the feminist movement.

Quiz: Conditional Probabilities

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

- $P(+x \mid +y) ?$

- $P(-x \mid +y) ?$

- $P(-y \mid +x) ?$

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W|T)$

$P(W|T = \text{hot})$

W	P
sun	0.8
rain	0.2

$P(W|T = \text{cold})$

W	P
sun	0.4
rain	0.6

Joint Distribution

$$P(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Calculating Conditional Distribution

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$\begin{aligned}P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\&= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\&= \frac{0.2}{0.2 + 0.3} = 0.4\end{aligned}$$



$$\begin{aligned}P(W = r|T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\&= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\&= \frac{0.3}{0.2 + 0.3} = 0.6\end{aligned}$$

$P(W|T = c)$

W	P
sun	0.4
rain	0.6

Normalization Trick

$$\begin{aligned}
 P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\
 &= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\
 &= \frac{0.2}{0.2 + 0.3} = 0.4
 \end{aligned}$$

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence



$P(c, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one)



$P(W|T = c)$

W	P
sun	0.4
rain	0.6

$$\begin{aligned}
 P(W = r|T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\
 &= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\
 &= \frac{0.3}{0.2 + 0.3} = 0.6
 \end{aligned}$$

Normalization Trick

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence



$P(c, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one)



$P(W|T = c)$

W	P
sun	0.4
rain	0.6

- Why does this work? Sum of selection is $P(\text{evidence})!$ ($P(T=c)$, here)

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}$$

Quiz: Normalization Trick

- $P(X | Y=-y)$?

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

SELECT the joint probabilities matching the evidence



NORMALIZE the selection (make it sum to one)



To Normalize

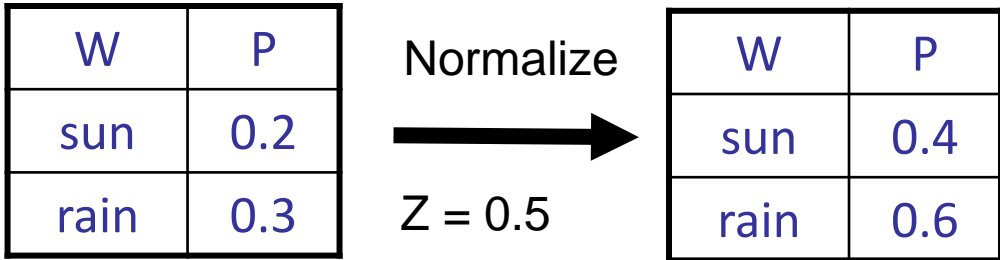
- (Dictionary) To bring or restore to a **normal condition**

All entries sum to ONE

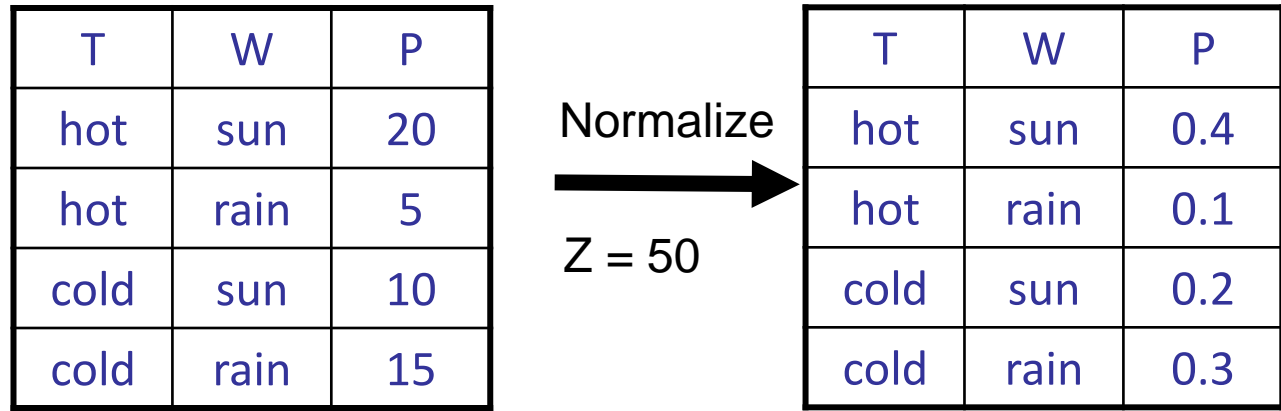
- Procedure:

- Step 1: Compute $Z = \text{sum over all entries}$
- Step 2: Divide every entry by Z

- Example 1



- Example 2



Normalization Trick

- A trick to get a whole conditional distribution at once:
 - Select the joint probabilities matching the evidence
 - Normalize the selection (make it sum to one)

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

→
Select

$P(T, r)$

T	R	P
hot	rain	0.1
cold	rain	0.3

→
Normalize

$P(T|r)$

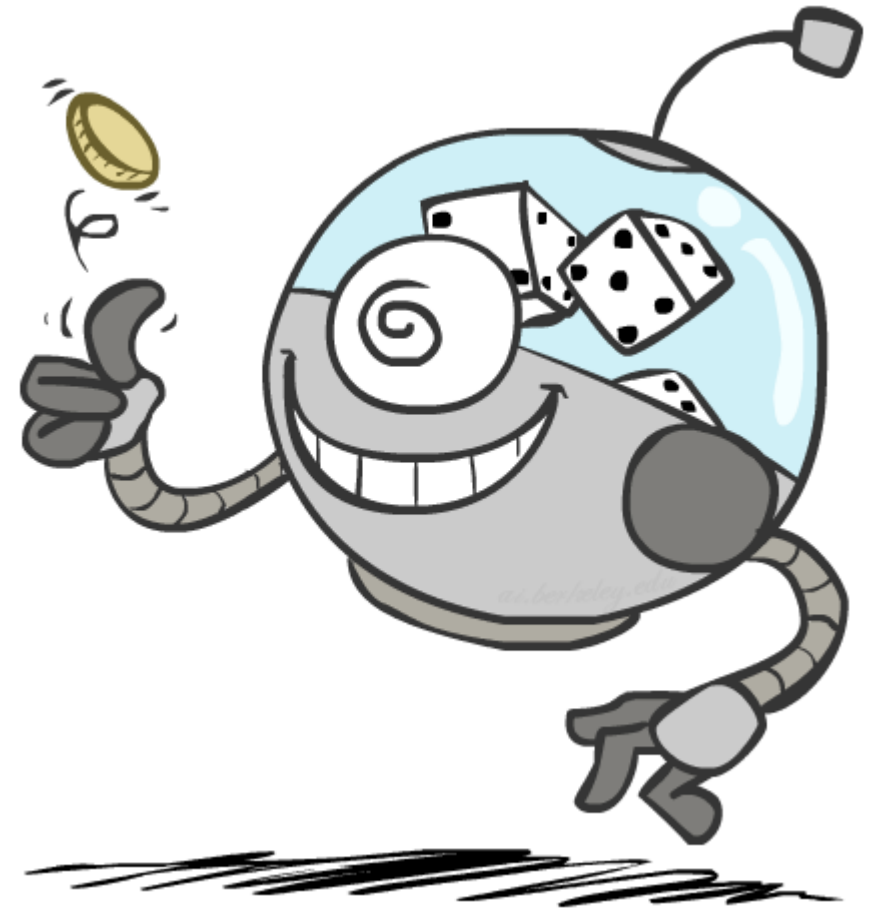
T	P
hot	0.25
cold	0.75

- Why does this work? Sum of selection is $P(\text{evidence})!$ ($P(r)$, here)

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}$$

Where we are...

- Probability
 - Random Variables
 - Joint and Marginal Distributions
 - Conditional Distribution
 - Product Rule, Chain Rule, Bayes' Rule
 - Inference
 - Independence



Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)
- We generally compute conditional probabilities
 - $P(\text{on time} \mid \text{no reported accidents}) = 0.90$
 - These represent the agent's *beliefs* given the evidence
- Probabilities change with new evidence:
 - $P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95$
 - $P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80$
 - Observing new evidence causes *beliefs to be updated*



Inference by Enumeration

- General case:

- Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
 - Query* variable: Q
 - Hidden variables: $H_1 \dots H_r$
- $\left. \begin{array}{l} X_1, X_2, \dots, X_n \\ \text{All variables} \end{array} \right\}$

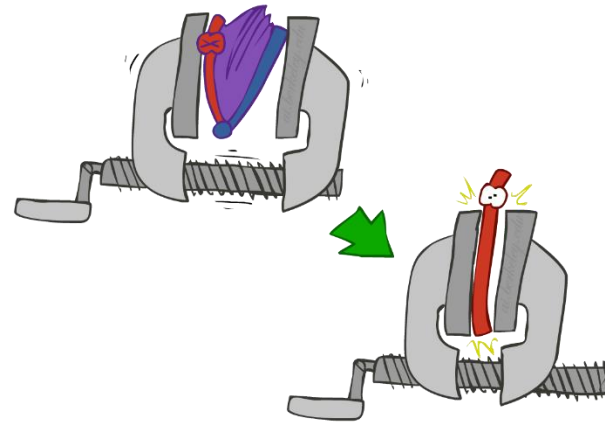
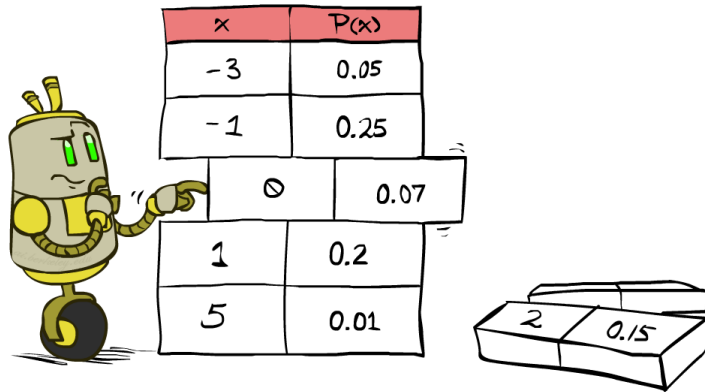
- We want:

$$P(Q|e_1 \dots e_k)$$

- Step 1: Select the entries consistent with the evidence

- Step 2: Sum out H to get joint of Query and evidence

- Step 3: Normalize



$$\times \frac{1}{Z}$$

$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} P(Q, \underbrace{h_1 \dots h_r, e_1 \dots e_k}_{X_1^{GWU}, X_2, \dots, X_n})$$

$$Z = \sum_q P(Q, e_1 \dots e_k)$$

$$P(Q|e_1 \dots e_k) = \frac{1}{Z} P(Q, e_1 \dots e_k)$$

Inference by Enumeration

- $P(W)$?
- $P(W \mid \text{winter})$?
- $P(W \mid \text{winter, hot})$?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- Obvious problems:
 - Worst-case time complexity $O(d^n)$
 - Space complexity $O(d^n)$ to store the joint distribution

The Product Rule

- Sometimes have conditional distributions but want the joint

$$P(y)P(x|y) = P(x, y) \quad \longleftrightarrow \quad P(x|y) = \frac{P(x, y)}{P(y)}$$

The Product Rule

$$P(y)P(x|y) = P(x, y)$$

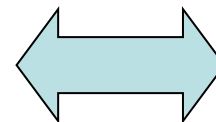
- Example:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3



$P(D, W)$

D	W	P
wet	sun	
dry	sun	
wet	rain	
dry	rain	---

The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions

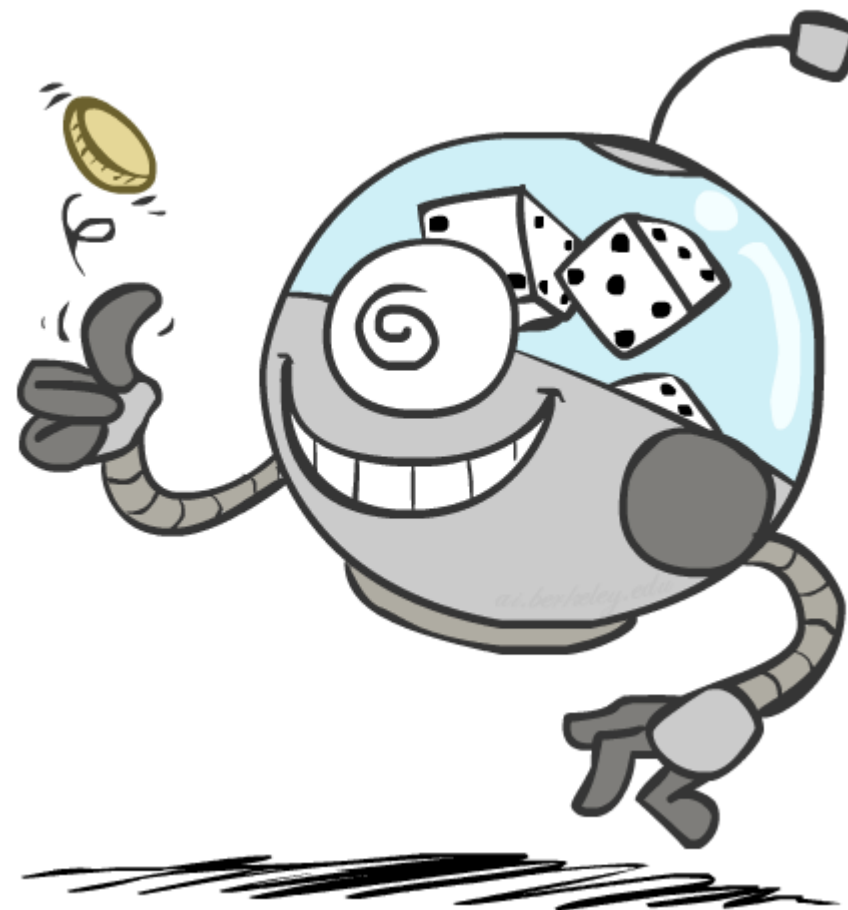
$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i|x_1 \dots x_{i-1})$$

- Why is this always true?

Where we are...

- Probability
 - Random Variables
 - Joint and Marginal Distributions
 - Conditional Distribution
 - Product Rule, Chain Rule, Bayes' Rule
 - Inference
 - Independence



Probabilities of Symptoms and Causes

BAYES RULE

Bayes Rule

In the running for
the most important
AI equation!

- Two ways to factor a joint distribution over two variables:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

- Dividing, we get:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

- Why is this important, or even helpful?
 - Allows us to build one conditional from its reverse
 - Often one conditional is tricky but the other one is simple
 - Many times, we know the probability of a symptom, but we need the probability of the underlying cause
 - Foundation of many systems



Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- Example:

- M: meningitis, S: stiff neck

$$\left. \begin{aligned} P(+m) &= 0.0001 \\ P(+s|+m) &= 0.8 \\ P(+s|-m) &= 0.01 \end{aligned} \right\} \text{Example givens}$$

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}$$

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?

Quiz: Bayes' Rule

- Given:

$$P(W)$$

R	P
sun	0.8
rain	0.2

$$P(D|W)$$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

- What is $P(W | \text{dry})$?
 - $P(\text{sun}/\text{dry}) = ?$
 - $P(\text{rain}/\text{dry}) = ?$

Ghostbusters

- Let's say we have two distributions:
 - Prior distribution** over ghost location: $P(G)$
 - Let's say this is uniform
 - Sensor reading model: $P(R | G)$
 - Given: we know what our sensors do
 - R = reading color measured at $(1,1)$
 - E.g. $P(R = \text{yellow} | G=(1,1)) = 0.1$
- We can calculate the **posterior distribution** $P(G | r)$ over ghost locations given a reading using Bayes' rule:

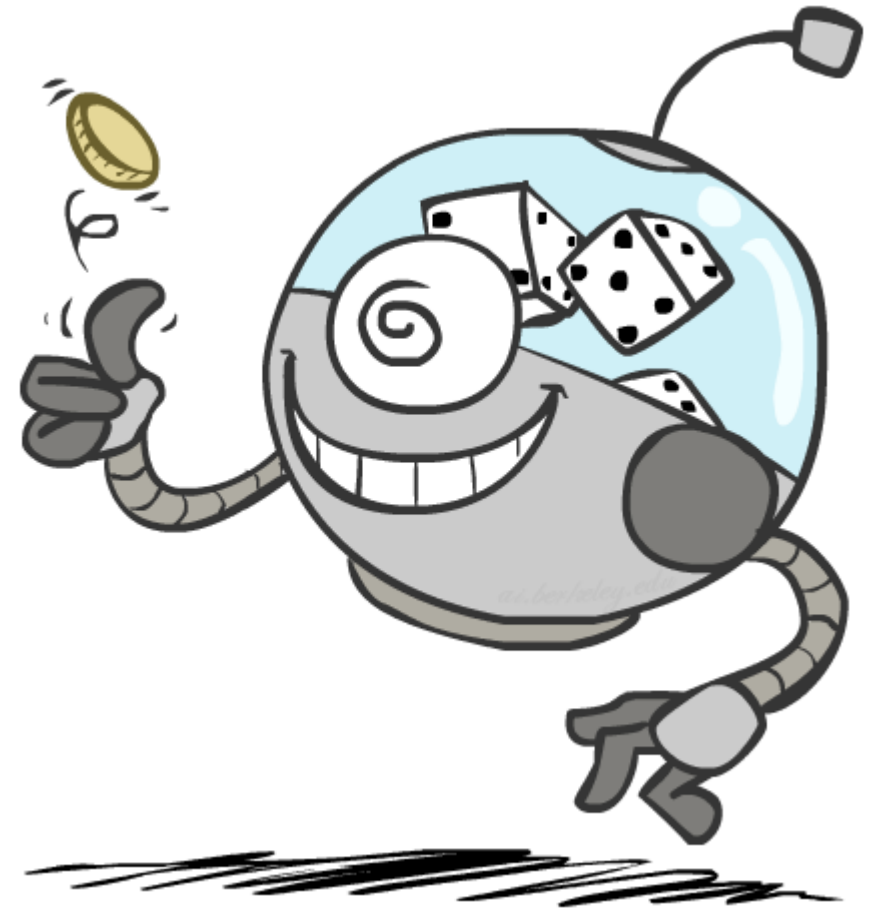
$$P(g|r) \propto P(r|g)P(g)$$

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

Where we are...

- Probability
 - Random Variables
 - Joint and Marginal Distributions
 - Conditional Distribution
 - Product Rule, Chain Rule, Bayes' Rule
 - Inference
 - Independence



Independence

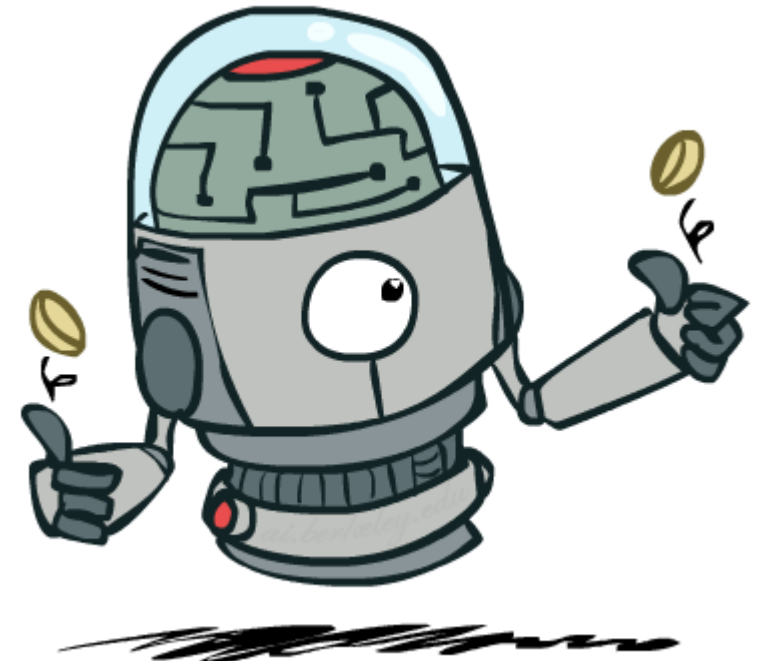
- Two variables are *independent* in a joint distribution if:

$$P(X, Y) = P(X)P(Y)$$

$$X \perp\!\!\!\perp Y$$

$$\forall x, y P(x, y) = P(x)P(y)$$

- Says the joint distribution *factors* into a product of two simple ones
 - Usually variables aren't independent!
- Can use independence as a *modeling assumption*
 - Independence can be a simplifying assumption
 - Empirical* joint distributions: at best "close" to independent
 - What could we assume for {Weather, Traffic, Cavity}?



Example: Independence?

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(T)$

T	P
hot	0.5
cold	0.5

$$P_2(T, W) = P(T)P(W)$$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

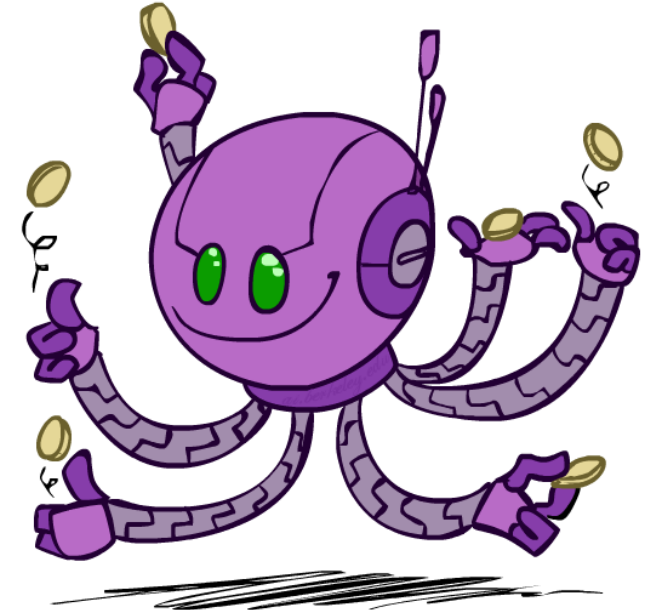
$P(X_2)$

H	0.5
T	0.5

...

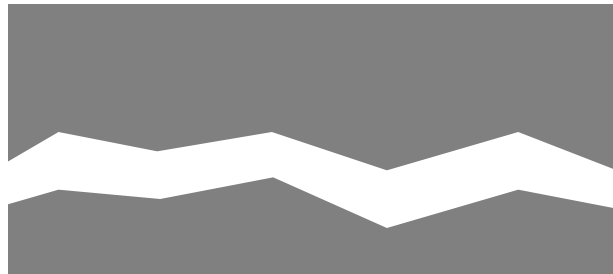
$P(X_n)$

H	0.5
T	0.5



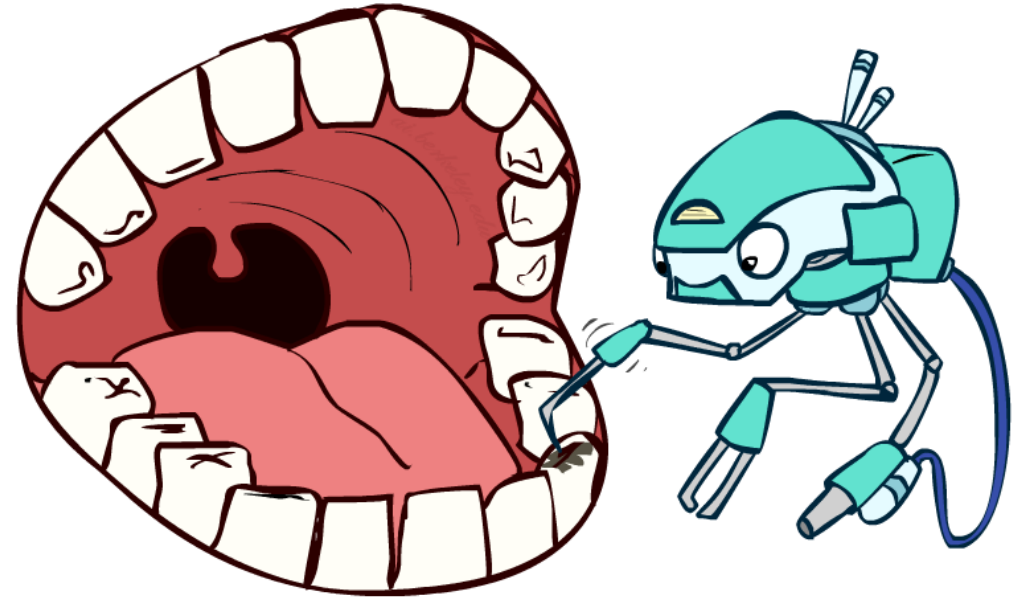
$P(X_1, X_2, \dots, X_n)$

2^n



Conditional Independence

- $P(\text{Toothache}, \text{Cavity}, \text{Catch})$
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
 - $P(+\text{catch} \mid +\text{toothache}, +\text{cavity}) = P(+\text{catch} \mid +\text{cavity})$
- The same independence holds if I don't have a cavity:
 - $P(+\text{catch} \mid +\text{toothache}, -\text{cavity}) = P(+\text{catch} \mid -\text{cavity})$
- *Catch is conditionally independent of Toothache given Cavity:*
 - $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$
- Equivalent statements:
 - $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
 - $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$
 - One can be derived from the other easily



Conditional Independence

- Unconditional (absolute) independence very rare
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z

$$X \perp\!\!\!\perp Y | Z$$

if and only if:

$$\forall x, y, z : P(x, y | z) = P(x | z)P(y | z)$$

or, equivalently, if and only if

$$\forall x, y, z : P(x | z, y) = P(x | z)$$

Conditional Independence

- What about this domain:

- Traffic
- Umbrella
- Raining

- What about this domain:

- Fire
- Smoke
- Alarm

Probability Recap

- Conditional probability $P(x|y) = \frac{P(x, y)}{P(y)}$
- Product rule $P(x, y) = P(x|y)P(y)$
- Chain rule
$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots \\ &= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $\forall x, y : P(x, y) = P(x)P(y)$
- X and Y are conditionally independent given Z if and only if: $X \perp\!\!\!\perp Y | Z$
 $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$

Summary

- Probability
 - Random Variables
 - Joint and Marginal Distributions
 - Conditional Distribution
 - Product Rule, Chain Rule, Bayes' Rule
 - Inference
 - Independence

